LIKELIHOOD AND LEAST-SQUARES APPROACHES TO THE M-CORNERED HAT

Charles A. Greenhall
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, California 91109

ABSTRACT

A simple model of the m-cornered hat estimation problem is set up and solved by the method of maximum likelihood. The method is compared by simulation to a least-squares method of Barnes and is shown to be inferior to it on the basis of mean square error. A bootstrap method of computing estimator performance is presented.

INTRODUCTION

Because fluctuations of frequency sources can be measured only by pairwise comparisons, the estimation of the noise level of each individual source is not straightforward. In the m-cornered hat problem there are m sources ($m \geq 3$); let the phase of the i^{th} source as function of time t be $\phi_i(t)$. The observations consist of the $m-1$ pair-differences $\phi_i(t) - \phi_1(t)$ over some stretch of time, and it is required to estimate the Allan variances, $\sigma_i^2(\tau)$, $i = 1$ to m, of all the sources, for some fixed τ.

We shall set up an oversimplified model of the situation, and show (without proof) how the unknown corner Allan variances can be estimated by the method of maximum likelihood. Using simulation, we shall compare the performance of these estimators to those generated by a weighted least squares approach of Barnes. Finally, a method for estimating the variances of the estimators themselves will be given.

A TOY MODEL

To attack the problem with a likelihood approach, a probabilistic model is needed. Here, the second phase differences of the i^{th} source, for a fixed τ, are represented by a set $x_i(t)$, $t = 1$ to n, of Gaussian random variables with mean zero and variance s_i. Thus, n is the number of samples of second phase difference. The crucial assumption is that all the mn random variables are independent.

The main reason for the toy status is that the second differences of phase can hardly ever be regarded as white. One can make a rough fit of the model to a more practical situation by letting n be the approximate degrees of freedom of the usual estimators of pair Allan variance [1][2][3,*Fig.*1]. This can be done in a rough way if the phase noise spectral densities of the sources are approximately proportional to each other in the vicinity of $f = 1/\tau$.

Since there are too many "variances" and "sigmas" in this field, the parameters s_i will henceforth be called "noise levels," a term also applied to the quantities s_{ij} defined below. The term "variance" will be reserved either for the theoretical variance of a random variable involved in the estimation process or for the sample variance of a finite sequence of numbers.

THE LIKELIHOOD FUNCTION

The observations now consist of the n-vectors $x_i - x_1$, for $i = 2$ to m, and it is required to estimate the parameter m-vector $s = (s_1, \ldots, s_m)$. The likelihood function is the probability density p of the observations given the parameters, regarded as a function of the parameters.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
DEC 1987		**00-00-1987 to 00-00-1987**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Likelihood and Least-Squares Approaches to the M-Cornered Hat	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **California Institute of Technology,Jet Propulsion Laboratory,4800 Oak Grove Drive,Pasadena,CA,91109**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Proceedings of the Nineteenth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Redondo Beach, CA, 1-3 Dec 1987

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

It is convenient to work with the object function

$$L = (-2 \log p)/n,$$

whose natural domain is the set of s such that all s_i are nonnegative with at most one s_i equal to zero. Thus, a point in the domain is either in the interior (all $s_i > 0$) or on the k^{th} wall ($s_k = 0$, other $s_i > 0$) for some k. Negative noise levels are not allowed.

For $i, j = 1$ to m define the observed pair noise levels by

$$s_{ij} = \frac{1}{n} \sum_{t=1}^{n} (x_i(t) - x_j(t))^2.$$

(Note $x_i - x_j = x_i - x_1 - (x_j - x_1)$.)

These correspond to practical estimates of pair Allan noise levels. Then the function L, with an additive constant neglected, is given by the formulas

$$L = \log(P/b) + Wb \qquad \text{(in the interior)}$$

$$= \log P_k + \sum_{j \neq k} \frac{s_{kj}}{s_j} \qquad \text{(on the } k^{th} \text{ wall)},$$

where

$$b = \frac{1}{\sum_i \frac{1}{s_i}},$$

$$W = \frac{1}{2} \sum_i \sum_j \frac{s_{ij}}{s_i s_j},$$

$$P = \prod_i s_i, \quad P_k = \prod_{i \neq k} s_i.$$

It is gratifying that the likelihood function depends on the observation vectors only through the pair noise levels s_{ij}. Notice that the special role played by $i = 1$ has disappeared. Tryon and Jones [4] also computed the likelihood function for this model, but in a nonsymmetric form.

MAXIMUM LIKELIHOOD SOLUTION

It is required to find the s that minimizes L. The function L is continuous on its domain (interior + walls), and it can be shown that a minimum exists. The author has not been able to prove that the minimum is unique, nor that any interior stationary point of L is a minimum. Nevertheless, we shall proceed on the basis of these assumptions, which seem to be valid in practice. Setting the partial derivatives of L to zero, and making other transformations that improve iteration performance, we find that an interior point s is a stationary point of L if and only if it satisfies the equations

$$s_i = b_i \left[\sum_j \frac{s_{ij}}{s_j} - \frac{m-1}{m-2} W_i b_i \right], \quad i = 1 \text{ to } m, \tag{1}$$

where

$$b_i = \frac{1}{\sum_{j \neq i} \frac{1}{s_j}}, \quad W_i = \frac{1}{2} \sum_{j \neq i} \sum_{k \neq i} \frac{s_{jk}}{s_j s_k}.$$

It can be shown that the best wall point, the point s that minimizes L on the walls, is given by

$$s_i = s_{k_b,i}, \quad i = 1 \text{ to } m,$$

where k_b is the index k that minimizes the product of s_{ki} for $i \neq k$. (The author does not know what happens if two of these products are equal.)

An algorithm for finding the maximum likelihood solution can now be given.

1. Let s be the best wall point.
2. Iterate equations (1) once to obtain a new s.
3. If s is an interior point then

   iterate equations (1) to convergence. The resulting interior point is the solution.

   Otherwise

   the best wall point is the solution.

The author can prove that the condition in step 3 is sufficient for an interior minimum of L, but has not been able to prove its necessity. Only experience has shown that if the best wall point does not iterate into the interior, then the best wall point is the minimum of L. Usually, about 10 iterations are adequate for the interior case. Occasionally, though, more than 100 are needed; the author does not know whether this is intrinsic to eq.(1) or caused by roundoff error.

A wall solution $s_k = 0$ does not mean that we believe the k^{th} noise level to be zero; it implies only that an interval of uncertainty for it goes from zero out to some positive value. Unfortunately, the author does not know how to generate confidence intervals for this method.

THE CLASSICAL 3-CORNERED HAT

Although it is not obvious, for $m = 3$ the set of equations (1) is indeed equivalent to the classical 3-cornered hat equations

$$s_i + s_j = s_{ij}, \quad i < j; \tag{2}$$

moreover, the maximum likelihood occurs on the k^{th} wall if and only if (2) has a nonpositive solution for s_k. Thus, for $m = 3$ one solves the three equations (2) in the usual way. If all s_i are positive, then s is the maximum likelihood solution. If s_1, say, is not positive, then the maximum likelihood solution is $s_1 = 0$, $s_2 = s_{12}$, $s_3 = s_{13}$.

A WEIGHTED LEAST-SQUARES APPROACH

If $m > 3$, the $m(m-1)/2$ equations (2) in m unknowns are overdetermined; Barnes [5] has suggested that estimators for the corner noise levels s_i might be obtained from a least-squares solution of this system. Since s_{ij} is proportional to a chi-squared variable with n degrees of freedom, it is reasonable to assume that the residual of the (i,j) equation has a standard deviation proportional to s_{ij} itself. In addition, the equations are treated as if their residuals were orthogonal, whereas in reality they are correlated in some unknown way.

Under these assumptions, Barnes showed how to compute the least-squares solution by a Kalman filter iteration, starting from some initial estimate of the solution. Here, we use

another solution method: the weighted system is reduced to an unweighted system simply by dividing each equation by s_{ij}, resulting in

$$\frac{s_i + s_j}{s_{ij}} = 1, \quad i < j. \tag{3}$$

This system can be solved with an algorithm of Lawson and Hanson[6] called NNLS (nonnegative least squares), which minimizes the sum of squares of the residuals over the set of s with nonnegative components. This has two advantages over the Kalman method of solution. First, no prior estimates of the solution are required; second, the Kalman solution (or unrestricted least squares) can produce negative s_i. A disadvantage is that NNLS produces no covariances for the estimators. We also point out that, even when $m = 3$, the wall solutions of NNLS are not the same as the maximum likelihood wall solutions.

SIMULATION RESULTS

Runs of 1000 trials of the toy model were made for various choices of true corner noise levels s_i. Here are some typical results.

Legend: ML = maximum likelihood,
NNLS = nonnegative least squares (weighted),
RMSE = root mean square error
= square root (bias^2 + sample variance).

m = 4 corners, 1000 simulation trials.

		n = 10 samples				20 samples			
		ML		NNLS		ML		NNLS	
i	True s_i	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	1.	.05	.94	.07	.82	.02	.66	.05	.62
2	2.	-.07	1.27	-.19	1.14	-.02	.91	-.04	.87
3	3.	.08	1.81	-.14	1.63	-.03	1.14	-.14	1.10
4	4.	-.08	2.13	-.36	2.01	-.04	1.46	-.26	1.41

The NNLS method has greater bias but less RMSE than the ML method. Only if the true s_i are much more unbalanced than the ones used here, say 1., 10., 10., 10., does ML show slightly better performance than NNLS. For a larger number of samples ($n = 100$), the RMSEs of the two methods are almost equal. On this basis, then, weighted least squares appears to be the method of choice.

Another example studies the effect of adding corners. One would hope that a 3-cornered hat estimate could be improved by adjoining another corner. This is so, at least on the basis of ensemble mean square error:

True variances = 1., n = 10 samples, 1000 simulation trials. RMSE averaged over m corners.

m	ML RMSE	NNLS RMSE
3	.66	.67
4	.62	.55
5	.59	.51
6	.57	.50

In this special case at least, there appears to be little benefit in going beyond 5 corners. The reader is cautioned that these results are valid only for averages over many trials; for an individual trial, the estimates of s_i can get **worse** as corners are added.

SECOND-MOMENT BOOTSTRAP OF ESTIMATE VARIANCES

This method, an adaptation of Efron's bootstrap method[7] to the present situation, is a Monte Carlo computation of the variance of an estimator of s_i, using nothing but a single set of observations. Suppose that one set of difference vectors $x_i - x_1$ is collected, the pair noise levels s_{ij} computed, and estimates of corner noise levels s_i computed by the ML or NNLS method. Form the matrix $R = (r_{ij},\ i,\ j = 2 \text{ to } m)$ by

$$r_{ij} = \frac{1}{2}(s_{1i} + s_{1j} - s_{ij}).$$

Then R is the matrix of inner products of the n-vectors $x_i - x_1$, hence is positive definite. (Indeed, this is the matrix that Tryon and Jones used for expressing the likelihood function, so that we are returning to their nonsymmetric formulation.) Let $Y_i(t)$, $i = 2$ to m, $t = 1$ to n, be zero-mean Gaussians such that 1) for each t, the $Y_i(t)$ have covariance matrix R; 2) if $s \neq t$, the random vectors $Y(s)$ and $Y(t)$ are independent. Also, let $Y_1(t) = 0$. Then

$$E(Y_i(t) - Y_j(t))^2 = s_{ij}.$$

This setup is called the bootstrap model. The random variables $Y_i(t)$ play a role similar to that of $x_i(t) - x_1(t)$ in the original toy model. To make a trial of the bootstrap model, one generates $m - 1$ independent unit Gaussians u_i, then expresses the $Y_i(t)$ as appropriate linear combinations of the u_i with coefficients obtained from the Choleski decomposition of R. This is repeated for $t = 1$ to n.

To use the model, one makes n_b (100 to 1000, say) computer simulation trials of it, always with the same observed s_{ij}. Each trial yields a bootstrap sample of the pair noise levels given by

$$S_{ij} = \frac{1}{n}\sum_{t=1}^{n}[Y_i(t) - Y_j(t)]^2,$$

from which a bootstrap sample of the ML or NNLS estimate of corner noise levels s_i is computed by the algorithms given previously. The sample variance of all the n_b bootstrap sample estimates of s_i is then the approximate result for the variance of the s_i estimate derived from the original s_{ij}.

This technique substitutes raw number-crunching power for the theoretical power to compute variances of estimators that are complicated functions of the observations. Here are two examples of its results. Note that "Est s_i" is the single ML or NNLS estimate whose variance we wish to estimate, and "Boot σ" is the square root of the sample variance of the bootstrap s_i estimates over 1000 trials of the bootstrap model. In each example, the original s_{ij} and estimates of s_i were obtained from a single trial of the toy model.

$m = 4$ corners, $n = 10$ samples, 1000 bootstrap trials

Example 1

		ML		NNLS	
i	True s_i	Est s_i	Boot σ	Est s_i	boot σ_i
1	1.	1.07	.72	.98	.58
2	2.	.69	.64	.75	.49
3	3.	1.57	.94	1.51	.73
4	4.	2.73	1.43	2.62	1.27

Example 2

		ML		NNLS	
i	True s_i	Est s_i	Boot σ	Est s_i	Boot σ
1	1.	2.01	1.08	1.98	1.00
2	2.	2.96	1.47	2.97	1.46
3	3.	4.36	2.06	4.38	2.04
4	4.	.21	.52	.22	.48

The only difference between these two examples is the seed for the random number generator that generated the original $x_i(t)$ from which the original s_{ij} were computed. The estimated s_i and true s_i differ by reasonably small multiples of boot σ, except for $i = 4$ of Example 2, which was particularly unlucky. It seems that if luck throws an estimate of s onto or close to one of the walls, then it is difficult to pull it away. The author does not know how to recognize this situation without prior knowledge of the true s_i.

For larger values of n, where (with high probability) the walls do not threaten, simpler methods of variance estimation can be used. For example, the Kalman filter or unrestricted least squares algorithm yields a covariance matrix of the estimators. As n grows, the bootstrap trials become more onerous anyway. To check the performance of the bootstrap for larger n, the following example for $n = 100$ compares the standard deviations produced by a bootstrap run with those produced by a regular simulation run of the toy model.

$m = 4$ corners, $n = 100$ samples, 1000 toy and bootstrap trials.

		ML		NNLS	
i	True s_i	Toy σ	Boot σ	Toy σ	Boot σ
1	1.	.29	.31	.29	.30
2	2.	.39	.42	.38	.42
3	3.	.53	.45	.52	.44
4	4.	.66	.61	.66	.59

The corresponding σ values agree within 15%, which indicates that the bootstrap estimates of variance are compatible with the true values and hence are compatible with those produced by more economical methods when n is large.

CONCLUSIONS

We have presented two methods for estimating the noise levels (Allan variances) of m frequency sources whose pair noise levels are measured. In most cases, the estimators produced by the weighted least squares method have smaller mean-square errors than the maximum likelihood estimators. (Nevertheless, the author is not yet willing to reject the maximum likelihood method out of hand.) Estimates of variance for the noise level estimates produced by either method can be obtained from a conceptually simple but computationally expensive bootstrap method whose results are not always reasonable.

REFERENCES

[1] P. Lesage and C. Audoin, Effect of dead time on the estimation of the two-sample variance, IEEE Trans. Instrum. Meas., vol. IM-28, pp.6-10, 1979.

[2] K. Yoshimura, Characterization of frequency stability: uncertainty due to the autocorrelation of the frequency fluctuations, IEEE Trans. Instrum. Meas., vol. IM-27, pp.1-7, 1978.

[3] C.A. Greenhall, A structure function representation theorem with applications to frequency stability estimation, IEEE Trans. Instrum. Meas., vol. IM-32, pp.364-370, 1983.

[4] P.V. Tryon and R.H. Jones, Maximum likelihood estimation of variances when the differences of Gaussian white noise series are observed, unpublished note received by the author from J.A. Barnes, no date.

[5] J.A. Barnes, Kalman decomposition of variances, unpublished note, 1982.

[6] C.L. Lawson and R.J. Hanson, Solving least squares problems, Prentice-Hall, 1974.

[7] B. Efron, The jackknife, the bootstrap, and other resampling plans, CBMS-NSF Regional Conference Series in Applied Mathematics 38, Society for Industrial and Applied Mathematics, 1982.

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Donald Percival, University of Washington: You are actually doing a constrained maximum likelihood, since you don't allow the negative variances. Have you done any experiments where you removed the constraints and did the full maximum likelihood solution? Perhaps that might improve the overall root mean square deviation.

Mr. Greenhall: The unknown parameters are essentially just the variances. I don't see how they could be made negative.

Mr. Percival: This problem occurs commonly in statistics in the analysis of random effects models. If you look at Chaffe's (sp.) book,on the analysis of variance, he deals with it. His solution is to use the model if it comes out negative. It does give you some information as to how large the negative thing is as to the quality of the model and so on. I am wondering if constraining the solution to lie in that region may somehow distort the root mean square criterion that you have. In other words, if one of the values were negative, the other two values might be a lot closer than to where they are, so the root mean square might be considerably improved.

Mr. Greenhall: I have not looked at the likelihood function outside of the non–negative region.

Mr. Percival: Just one other question. What was the model that you used in order to do the bootstrapping? You had to account for the correlations somehow.

Mr. Greenhall: You get these observed s_i's. Then you generate the m Gaussians such that the expected square between the two of those is equal to the s_{ij}'s. Then you make n independent copies of those. That is the bootstrap model.